Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/005715

International filing date: 17 February 2005 (17.02.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/635,397
Filing date: 10 December 2004 (10.12.2004)

Date of receipt at the International Bureau: 23 March 2005 (23.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1295810

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 14, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/635,397***
FILING DATE: ***December 10, 2004***
RELATED PCT APPLICATION NUMBER: ***PCT/US05/05715***

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PTO/SB/16 (12-04)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EV 333611944 US

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
RONEN EZRA	BASRI	REHOVOT, ISRAEL

Additional inventors are being named on the SECOND PAGE *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max):

MULTISCALE SEGMENTATION BY COMBINING MOTION AND INTENSITY CUES

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☑ The address corresponding to Customer Number: 27317

OR

☑ Firm or Individual Name FLEIT, KAIN, GIBBONS, GUTMAN, BONGINI & BIANCO, P.L.

Address 601 BRICKELL KEY DRIVE, SUITE 404

City MIAMI	State FL	Zip 33131
Country USA	Telephone 305-416-4490	Fax 305-416-4489

ENCLOSED APPLICATION PARTS *(check all that apply)*

□ Application Data Sheet. See 37 CFR 1.76

☑ Specification *Number of Pages* 16

☑ Drawing(s) *Number of Sheets* 2

□ CD(s), Number of CDs ______

□ Other (specify) ______

Application Size Fee: If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

METHOD OF PAYMENT OF FILING FEES AND APPLICATION SIZE FEE FOR THIS PROVISIONAL APPLICATION FOR PATENT

☑ Applicant claims small entity status. See 37 CFR 1.27.

□ A check or money order is enclosed to cover the filing fee and application size fee (if applicable).

☑ Payment by credit card. Form PTO-2038 is attached

□ The Director is hereby authorized to charge the filing fee and application size fee (if applicable) or credit any overpayment to Deposit Account Number: 500601. A duplicative copy of this form is enclosed for fee processing.

TOTAL FEE AMOUNT ($) $100.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No.

□ Yes, the name of the U.S. Government agency and the Government contract number are: ______

SIGNATURE ______ Date DECEMBER 10, 2004

TYPED or PRINTED NAME MARTIN FLEIT

REGISTRATION NO. 16,900
(if appropriate)

TELEPHONE 305-416-4490

Docket Number: 7040-X04-069P

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (12-04)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

First Named Inventor	RONEN EZRA BASRI	Docket Number	7040-X04-069P

INVENTOR(S)/APPLICANT(S)

Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
MEIRAV	GALUN	REHOVOT, ISRAEL
ALEXANDER	APARSTSIN	REHOVOT, ISRAEL

Number 2 of 2

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

PTO/SB/17 (12-04)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL For FY 2005

☑ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 100

Complete if Known	
Application Number	
Filing Date	DECEMBER 10, 2004
First Named Inventor	RONEN EZRA BASRI
Examiner Name	
Art Unit	
Attorney Docket No.	7040-X04-069P

METHOD OF PAYMENT (check all that apply)

□ Check ☑ Credit Card □ Money Order □ None □ Other (please identify):

☑ Deposit Account Deposit Account Number: 500601 Deposit Account Name: MARTIN FLEIT

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

□ Charge fee(s) indicated below □ Charge fee(s) indicated below, **except for the filing fee**

☑ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17 ☑ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES

	FILING FEES		SEARCH FEES		EXAMINATION FEES		
Application Type	Fee ($)	Small Entity Fee ($)	Fee ($)	Small Entity Fee ($)	Fee ($)	Small Entity Fee ($)	Fees Paid ($)
Utility	300	150	500	250	200	100	
Design	200	100	100	50	130	65	
Plant	200	100	300	150	160	80	
Reissue	300	150	500	250	600	300	
Provisional	200	100	0	0	0	0	100

2. EXCESS CLAIM FEES

Fee Description	Fee ($)	Small Entity Fee ($)
Each claim over 20 or, for Reissues, each claim over 20 and more than in the original patent	50	25
Each independent claim over 3 or, for Reissues, each independent claim more than in the original patent	200	100
Multiple dependent claims	360	180

Total Claims Extra Claims Fee ($) Fee Paid ($)
______ - 20 or HP = ______ x ______ = ______
HP = highest number of total claims paid for, if greater than 20

Multiple Dependent Claims
Fee ($) \_\_\_\_\_\_ Fee Paid ($) ______

Indep. Claims Extra Claims Fee ($) Fee Paid ($)
______ - 3 or HP = ______ x ______ = ______
HP = highest number of independent claims paid for, if greater than 3

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets Extra Sheets Number of each additional 50 or fraction thereof Fee ($) Fee Paid ($)
______ - 100 = ______ / 50 = ______ (round up to a whole number) x ______ = ______

4. OTHER FEE(S) Fees Paid ($)

Non-English Specification, $130 fee (no small entity discount) ______

Other: ______ ______

SUBMITTED BY			
Signature	Martin Fleit	Registration No. (Attorney/Agent) 16,900	Telephone 305-416-4490
Name (Print/Type)	MARTIN FLEIT		Date DECEMBER 10, 2004

This collection of information is required by 37 CFR 1.136. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PATENT Attorney Docket No.: 7040-X04-069P

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Ronen Ezra BASRI, Meirav GALUN and Alexander APARTSIN

Serial No:

Filed: December 10, 2004

Title: MULTISCALE SEGMENTATION COMBINING MOTION AND INTENSITY CUES

CERTIFICATE OF EXPRESS MAILING

PATENTS
EXPRESS "Express Mail" Mailing Label number EV 333611944 US
Date of Deposit December 10, 2004

I hereby certify that the attached paper(s) or fee(s) is/are being deposited with the United States Postal Services "Express Mail Post Office to Addressee" service under 37 CFR §1.10 on the date indicated above and is addressed to the Commissioner for Patents, P.O. Box 1450 Alexandria, Virginia 22313-1450.

Estefania Belaunde
(Signature of person mailing paper or fee)

ESTEFANIA BELAUNDE
(Typed or printed name of person mailing paper or fee)

Multiscale Segmentation by Combining Motion and Intensity Cues

Ronen Basri, Meirav Galun, Alexander Apartsin
Weizmann Institute of Science

Abstract

Motion provides a strong cue for segmentation. In this paper we present a multiscale method for motion segmentation. Our method begins with local, ambiguous optical flow measurements. It uses a process of aggregation to resolve the ambiguities and reach reliable estimates of the motion. In addition, as the process of aggregation proceeds and larger aggregates are identified it employs a progressively more complex model to describe the motion. In particular, we proceed by recovering translational motion at fine levels, through affine transformation at intermediate levels, to 3D motion (described by a fundamental matrix) at the coarsest levels. Finally, the method is integrated with a segmentation method that uses intensity cues. We further demonstrate the utility of the method on both random dot and real motion sequences.

1. Introduction

Segmentation of objects based on their motion is perceptually striking, as is exemplified by motion sequences containing random dots. Finding satisfactory algorithmic solutions to this problem, however, has remained a challenge. Algorithmic approaches to motion segmentation seem to face both the difficulties that complicate the task of intensity-based segmentation along with the challenges that make motion estimation hard. Issues that complicate segmentation include devising an appropriate measure of similarity and rules of clustering to correctly separate the various segments. Similarly, difficulties in motion estimation are due to the sparseness of motion cues, particularly their absence in uniform regions and due to the aperture problem. Furthermore, of crucial importance is the selection of an appropriate motion model.

A number of effective algorithms have been proposed to address the problem of motion segmentation, many of which produce convincing results on quite complex motion sequences. These algorithms differ in the kind of information they use (sparse features versus dense intensity information) and the motion model they impose (2D parametric versus motion in 3D). Some of these approaches also recognize the importance of combining optical flow measurements with intensity information to solve the problem of motion segmentation.

Motion segmentation approaches that use dense intensity information largely impose 2D parametric motion models (mostly translation or affine). These include layered representations [19,21] (see also [2], [1,20] attempt to relax some of the main requirements of layered approaches), variational methods [5,3], graph-cuts algorithms [6,12,15], and sequential dominant motion removal [13]. Handling 3D motion is usually achieved by extracting and tracking a sparse set of features. Among these are subspace methods, which are restricted to orthographic projection [4,8,11,20] ([22] attempt to apply these methods directly to intensities). Other feature-based methods deal also with perspective projection [16,17].

In this paper we describe a multiscale scheme that enables, through the use of hierarchical bottom-up processing, to overcome some of the crucial difficulties in motion segmentation. In particular, our scheme combines motion with intensity cues. The method determines the segments adaptively and estimates their motion by varying the motion model according to the amount of statistics that appears in each segment. In particular, we have implemented three motion models, translation, affine, and 3D rigid motion followed by perspective projection. The method we present is a two-frame approach developed to work with small motions. It uses the weighted aggregation procedure of [14] as a platform for segmentation. Finally, the method is efficient, with linear runtime complexity in the number of pixels. We further demonstrate the utility of our method through experiments on both random dot and real image pairs.

2. Aggregation of Motion Cues

In this section we focus our attention on motion cues alone and defer the discussion of how we integrate them with intensity-based segmentation to Section 3. The method we present performs motion segmentation by applying a sequence of *coarsening* steps, each of which further clusters the pixels in the image according to their motion, producing fewer aggregates of pixels of larger sizes. We refer to the aggregates produced by each coarsening step as a *level of scale*, with the *fine* levels containing the small aggregates produced after the first few steps of the algorithm, and the *coarse* levels containing the larger aggregates produced in subsequent steps. The primary objective of these coarsening steps is to determine which collections of pixels share a unified motion. This is achieved by simultaneously resolving motion ambiguities and describing the motion by the appropriate model. At this point we consider three types of motions - translation, affine, and rigid transformation in 3D. In the future we plan to extend this by handling 2D-homographies and nonrigid transformations.

Every coarsening step is composed of two parts, clustering and re-estimation. For clustering we select a set of *seed* elements, and then associate every element from the previous level to these seeds by soft assignment. Once the clusters are determined we estimate the common motion of the cluster. Here the parameters of the motion are determined and ambiguities are resolved.

As this iterative coarsening procedure proceeds we gradually modify the model used to describe the motion of aggregates. At finer levels we seek to determine the translation of aggregates. We achieve this by applying a process of sharpening the raw motion cues. This process allows us to identify either the translation of the center of mass of an aggregate or a 1-D constraint on this motion. Later on, as sufficient translational information is accumulated, we use this information to determine more complex motions, including affine transformation and rigid motion in 3D.

Below we describe the different components of the algorithm.

2.1 Initial Optical Flow Measurements

Measuring optical flow is complex, partly because local information usually is insufficient to determine the motion of a given pixel. In particular, pixels near edges are subject to a 1-D aperture ambiguity, and pixels within uniform regions are subject to a 2-D ambiguity. To represent this ambiguity we chose to follow the method of [15] and represent the initial optical flow measurements as a *motion profile*.

Let Im_1 and Im_2 denote the two input images. Using homogeneous coordinates, let $\mathbf{x}_i = (x_i, y_i, 1)^T$ denote a pixel in Im_1. The motion profile $M_i(\mathbf{u})$ is a normalized 2-D histogram reflecting our estimate of the probability that the optical flow at $\mathbf{x}_i$ is given by $\mathbf{u} = (u, v, 0)^T$. To estimate this histogram we compare a 3×3 window from Im_1 centered at $\mathbf{x}_i$ with similar windows in Im_2 centered at offsets $\mathbf{u}$ within a bounded distance from $\mathbf{x}_i$. Using the SSD (sum of squares distance) between the intensities in the two windows we set the motion profile to be

$$M_i(\mathbf{u}) = \frac{1}{Z}\left(e^{-\alpha SSD(\mathrm{Im}_1(\mathbf{x}_i), \mathrm{Im}_2(\mathbf{x}_i + \mathbf{u}))} + C\right). \qquad (1)$$

The constant α controls the penalty due to difference in intensity; assuming brightness constancy, α should be set according to the level of noise in the images. The constant term C is added to ensure that no offset is assigned zero probability (since there is always a chance that the pixel changes its intensities after motion, e.g., due to occlusion). Finally, Z is a normalizing factor set to ensure that the entries in the histogram sum to 1. We can in general use a prior to modulate this expression (e.g., incorporating small motion assumption). In our implementation we used a uniform prior, resulting in the expression (1). Another issue is how to initialize the motion profile of pixels near the boundaries of the image. Denote by k the number of cells contained in a profile, we assign $1/k$ to each cell corresponding to motion that exceeds the boundaries of the image. The rest of the cells are assigned proportionally to the expression (1).

2. 2 Optical Flow Disambiguation

To handle translation, we make the simplifying assumption that pixels provide independent information about their motion and use this assumption to evaluate which pixels should cluster together. According to this independence assumption, the joint probability that two pixels $\mathbf{x}_i$ and $\mathbf{x}_j$ with motion profiles M_i and M_j share a common translation $\mathbf{u}$ is given by $M_i(\mathbf{u})M_j(\mathbf{u})$. In reality, we may want to cluster together neighboring pixels that share a common motion even if this motion is non-translational (e.g., rotation). To account for small deviations from translation, and to account for noise, we first smooth the two motion profiles before taking their product. Denote by $g(M(\mathbf{u})) = G * M(\mathbf{u})$, where G denotes a Gaussian function with zero mean and a small standard deviation σ (we used $\sigma = 0.5$), and '$*$' denotes convolution. Then the chance that two pixels share roughly the same translation is given by

$$g(M_i(\mathbf{u}))g(M_j(\mathbf{u})). \qquad (2)$$

To evaluate the resemblance of the motion profiles of two neighboring pixels we follow [15] and define a measure based on the normalized correlation between the profiles. Define the similarity between two profiles as $d_{\text{profile}} = 1 - g(M_i(\mathbf{u}))g(M_j(\mathbf{u}))$ and

$$w_{ij} = e^{-\beta d_{\text{profile}}},(3)$$

where β is a scaling factor. At the finest level each pixel is connected to its four immediate neighbors with $w_{ij} > 0$, and every furthest pairs satisfy $w_{ij} = 0$.

Each coarsening step begins by selecting a subset of the elements from the previous level (pixels in the finest level, aggregates of pixels in higher levels) as *seeds*, with the constraint that all other elements are strongly associated with (subsets of) these seeds (using the similarity in (3) as a measure of association). We further prescribe an association weight p_{ik} to express the strength of association of a finer element i to a coarser seed k:

$$p_{ik} = \frac{w_{ik}}{\sum_l w_{il}}, \qquad (4)$$

where the denominator is summed over all seeds. Note the values p_{ik} are positive only in a close neighborhood of each element.

Once all the association weights are determined we can construct a common motion profile for the new aggregates. By generalizing (2) we obtain

$$M_k(\mathbf{u}) = \frac{1}{Z}\prod_i g(M_i(\mathbf{u}))^{p_{ik}V_i/\bar{V}}. \qquad (5)$$

According to this expression the motion profile of an aggregate k is given by the product of all the motion profiles of its children, where the power weighs each term according to the strength of association of a child to the seed and accounts for its volume (with V_i the volume of child i and $\bar{V}$ the average volume of an aggregate in its level). Z is the appropriate normalizing constant. With this formula the motion profile of a pixel is distributed between all the seed it is associated with.

This coarsening process, which is composed of seed selection, associating elements with the seeds, and computing new motion profiles, is repeated creating at every subsequent level fewer aggregates of larger sizes. Expressing the motion profile of these aggregates as a product of the profiles of their children results in a sharp decrease of the probabilities of incompatible motions relative to that of the correct translation. In textured regions it is often sufficient to perform one or two coarsening steps to obtain a sharply peaked motion profile. In contrast, motion profiles within uniform regions usually remain ambiguous until either the aggregation process combines them with areas that contain features or the entire uniform region is clustered. Combining this process with intensity based segmentation (Section 1) ensures the clustering of both uniform regions and texture features, and this in turn assists in obtaining meaningful optical flow measurements.

During this aggregation process we examine the motion profile of each of the aggregates in each level to determine if its profile is sharply peaked, in which case we label the aggregate as *peaked* and set the optical flow for the center of the aggregate according to the location of the peak. If a profile is not sharply peaked we further examine it to test whether it contains a line with elevated probability values. In that case we conclude that the aggregate is subject to an aperture problem. We thus label the aggregate as *bar-peaked* and associate a normal flow vector to the center of the aggregate according to the parameters of this line.

2.3 Affine Transformation

As we proceed with the aggregation process, the size of aggregates increases, and translation ceases to accurately reflect their motion. We therefore wish to use a more complex model to describe this motion. We do so by fitting an affine transformation for each aggregate. Unfortunately, the motion profile of an aggregate cannot be used to determine its affine motion since it does not contain sufficient degrees of freedom. Instead, we accumulate constraints from the peaked and bar-peaked sub-aggregates, and use these constraints to determine the affine motion.

Specifically, suppose the flow at a point $\mathbf{x}_i$ is given by $\mathbf{u}$. We wish to fit a 3×3 matrix A that satisfies $A\mathbf{x}_i = \mathbf{u}$ (with the last row of A containing $(0,0,0)$). There are two weights we need to take into account. The first is the degree to which $\mathbf{x}_i$ (which denotes the center of mass of some sub-aggregate i a few levels down) belongs to the aggregate k for which we perform the computation. Generalizing (4) to i and k separated by any number of levels, we denote this weight by p_{ik}. The second weight reflects our belief in the optical flow measurement, as is expressed in the motion profile, given by $M_i(\mathbf{u})$. Incorporating these weights we seek a matrix A that minimizes

$$\min_A \sum_i \sum_{\mathbf{u}} p_{ik} M_i(\mathbf{u}) \left\| A\mathbf{x}_i - \mathbf{u}_i \right\|^2, \tag{6}$$

with the summation going over all sub-aggregates i and their motion profiles $\mathbf{u}$. Taking derivatives with respect to A we obtain[1]

$$A\left(\sum_i p_{ik}\mathbf{x}_i\mathbf{x}_i^T\right) = \sum_i \sum_{\mathbf{u}} p_{ik} M_i(\mathbf{u})\,\mathbf{u}\mathbf{x}_i^T. \tag{7}$$

This provides a set of six equations in the six unknown components of A, which uses moments of points up to second order (left hand side) and bilinear moments of their motion (right hand side). We collect these moments from the peaked and bar-peaked sub-aggregates of all finer levels using their motion profiles.

To obtain these moments we apply a process of accumulation in which we

[1] This and subsequent derivations can be obtained using $\partial \mathbf{y}^T A\mathbf{x}/\partial A = \mathbf{y}\mathbf{x}^T$, and $\partial \mathbf{y}^T A^T A\mathbf{x}/\partial A = A(\mathbf{y}\mathbf{x}^T + \mathbf{x}\mathbf{y}^T)$.

use the moments computed at every level to compute the moments of the subsequent level. A straightforward accumulation of moments may result in bias, as the motion profile can suffer from noise or the motion profile may still be ambiguous. We therefore apply a *selective* moment aggregation in a way that only peaked or bar-peaked sub-aggregates contribute to the moment accumulation. We label an aggregate as peaked (or bar-peaked) *by heredity* if at least one of its strongly-related children is labelled peaked. In this case we compute its moments as a weighted sum of the moments of its peaked (or bar-peaked) children. If an aggregate is not labeled peaked (or bar-peaked) by heredity we further examine if most of the energy in its motion profile is concentrated around a single location (or a line), in which case we label the aggregate as peaked (respectively bar-peaked) and initialize its moments using the following expression:

$$\sum_{\mathbf{u}} M_k(\mathbf{u}) x^{\delta_1} y^{\delta_2} u^{\delta_3} v^{\delta_4}, \tag{8}$$

where (x,y) are the center of mass of the aggregate, $\delta_j \geq 0$ are integers and $\sum \delta_j \leq 2$. Note that the moments accumulated this way adaptively collect information from aggregates of different scales.

The zero order moment indicates the number of points contributing to the moments. Since a peaked aggregate contributes two equations and a bar-peaked contributes one equation, we can use the zero order moment to determine if a sufficient number of points have been identified to determine an affine transformation. Whenever we detect aggregates for which there are no sufficient constraints to determine an affine transformation we assign to them the identity matrix for the linear part and translation according to the constraints available. If no constraints are available we consider these aggregates as stationary.

Once we describe the motion of aggregates by an affine transformation further coarsening requires us to compare these motions. A simple way to compare affine transformations is by directly comparing their components. However, a significant difference in the components of a transformation may not necessarily imply a similar difference in the effect of the transformation. To account for this we compare two affine transformations by the difference between the motions they induce on the relevant aggregates. We use a top-down process in which we examine sub-aggregates at two finer levels down. Denote by A_k the affine transformation of aggregate k, and the center of mass of its sub-aggregates (two levels down) by $\mathbf{x}_i = (x_i, y_i, 1)^T$ and their respective association weights to k by p_{ik}. The (non-symmetric) difference between the affine transformations assigned to aggregates k and l, d_{kl}, is defined as

$$d_{kl} = \left(\frac{\sum_i p_{ik}((I+A_k)\mathbf{x}_i - (I+A_l)\mathbf{x}_i)^2}{\sum_i p_{ik}} \right)^{\frac{1}{2}}. \tag{9}$$

Similarly, the difference in the other direction d_{lk} is calculated. The joint affine transformation distance between the aggregates k and l is

$$d_{\text{affine}} = \min(d_{kl}, d_{lk}). \quad (10)$$

From a certain level on we substitute d_{profile} in (3) by this measure.

2.4 Fundamental Matrix

Video sequences are often taken with the camera moving. Generically, such a motion produces perspective distortions throughout the image, making it difficult to separate the moving objects from a stationary background. To account for this we compute for each aggregate at the top-most levels a fundamental matrix and compare the obtained matrices in an attempt to cluster together segments describing stationary portions of the scene. Below we describe how this fundamental matrix is computed and compared.

Using the same notation introduced in the previous section we seek a 3×3 rank 2 matrix F that minimizes

$$\min_F \sum_i \sum_{\mathbf{u}} p_{ik} M_i(\mathbf{u})((\mathbf{x}_i + \mathbf{u})^T F \mathbf{x}_i)^2. \quad (11)$$

Taking derivatives with respect to F we obtain

$$\sum_i \sum_{\mathbf{u}} p_{ik} M_i(\mathbf{u})((\mathbf{x}_i + \mathbf{u})^T F \mathbf{x}_i)(\mathbf{x}_i + \mathbf{u})\mathbf{x}_i^T = 0. \quad (12)$$

This provides a set of nine homogeneous equations in the components of F, which uses motion moments of points up to fourth order (defined as in (8) with $0 \le \delta_j \le 4$). We collect these moments from the sharp sub-aggregates only. (The bar-peaked sub-aggregates generally do not constrain the fundamental matrix since the constraint line may intersect many epipolar constraints.) We solve the equation using the eight point algorithm using the normalization procedure proposed by Hartley [10] followed by rank reduction. Degeneracies can be handled as in [18], although this has not yet been implemented in our method.

The calculation of the fundamental matrix is followed by a comparison of the fundamental matrices between two neighboring aggregates. In this case it is not straightforward to apply the same comparison procedure used in the affine case, since a fundamental matrix provides only a line constraint on the location of each point. We therefore chose to use the simpler method of comparing the entries of the two matrices using an l_2 norm. The resulting measure, $d_{\text{fundamental}}$ is then use to replace d_{profile} in (3).

Finally, an outline of our motion segmentation algorithm is provided in Table 1.

Motion Segmentation:

- Given two images Im_1 and Im_2, prepare for each pixel in Im_1 a motion profile (1).
- Assign a weight to each pair of neighboring pixels according to the normalized correlation between their motion profiles (3).
- **Coarsening iteration**:
  1. **Clustering**:

(a) Select a set of seeds such that the remaining elements are strongly connected to this set.

(b) Define the strength of association of a fine element *i* to a coarse seed *k* (4).

2. **Re-estimation**: For each seed

   (a) Calculate the motion profile of the seed by multiplying the profiles of its children (5).

   (b) Examine whether the seed is peaked, by heredity or by itself.

   (c) If it is not peaked check if it is bar-peaked, by heredity or by itself.

   (d) Accumulate adaptively, moments (orders one to four) originated by peaked seeds.

   (e) Accumulate separately, moments (orders one and two) originated by bar-peaked seeds.

   (f) If there is enough statistics, calculate affine transformation by merging moments from peaked and bar-peaked profiles.

   (g) If there is enough statistics, calculate fundamental matrix from peaked profiles.

3. Calculate for each neighboring seeds cross correlation distance, affine transform distance and fundamental matrix distance.

4. Modify appropriately the similarities between neighboring seeds.

Table 1. Outline of the motion segmentation algorithm

3. Using Intensity Cues

To combine motion with intensity cues we integrate our motion segmentation algorithm with the Segmentation by Weighted Aggregation (SWA) algorithm [14]. This algorithm has been extended to also handle texture cues [7], although our implementation did not make use of these cues. Below we describe first the main principles behind the SWA algorithm (Section 1), and then we discuss how we combine motion with intensity cues in this framework (Section 2).

3.1. SWA Segmentation

The SWA algorithm is a multiscale graph partitioning algorithm. Given an image, it constructs a graph $G=(V,W)$, with nodes in V representing image pixels and the symmetric edge weight matrix W representing the affinities between neighboring pixels. To evaluate segments it defines a saliency measure as follows. Every node v_i, ($1 \le i \le N$, where $N = \|V\|$) is associated with a state variable u_i, and every candidate segment $S = \{v_1, v_2, ..., v_m\} \subseteq V$ is associated with a state $u = (u_1, u_2, ..., u_N)$ such that

$$u_i = \begin{cases} 1 & \text{if } v_i \in S \\ 0 & \text{if } v_i \notin S. \end{cases} \tag{13}$$

The saliency associated with S is defined by

$$\Gamma(S)=\frac{u^T Lu}{\frac{1}{2}u^T Wu}, \tag{14}$$

where L is the Laplacian matrix whose elements are

$$l_{ij}=\begin{cases}\sum_{k(k\neq i)} w_{ik} & i=j\\ -w_{ij} & i\neq j.\end{cases} \tag{15}$$

This saliency measure sums the weights along the boundaries of S normalized by the internal weights. Segments that yield small values of $\Gamma(S)$ are considered salient. Allowing arbitrary real assignments to u the minimum for Γ is obtained by the minimal generalized eigenvector u of

$$Lu=\lambda Wu. \tag{16}$$

with the condition that $\lambda>0$.

The SWA algorithm finds the best partitions (0-1 assignments of u by recursively producing smaller representative sets of seeds, such that every pixel is strongly-connected to the set of seeds. Denote by $U=(U_1,U_2,...,U_n)$ the coarse level state vector. We construct a sparse, $N\times n$ matrix P such that

$$u\approx PU. \tag{17}$$

P is called the inter-scale interpolation matrix. Using this matrix the saliency measure Γ can be written as

$$\Gamma=\frac{u^T Lu}{\frac{1}{2}u^T Wu}\approx\frac{U^T P^T LPU}{\frac{1}{2}U^T P^T WPU}. \tag{18}$$

To calculate the right hand side of this equation we need to compute the matrix $P^T WP$, which is a coarse representation of the original weight matrix. This is called *weighted aggregation*. Exploiting the sparseness of P, this product is computed in linear time.

3.2. Combining Motion with Intensities

In this section we will attach the superscripts M and I to denote measures corresponding to motion and intensity information respectively. The multiscale partitioning procedure described in the previous section can be used for segmentation combining motion and intensity cues in the following way. Given a pair of images, we begin by constructing a 4-connected graph $G=(V,W)$, where every pixel is represented by a node $v_i\in V$, and every pair of neighboring pixels are connected with an edge with weight w_{ij}. This weight is a product of two term. A measure reflecting the contrast between the two pixels i and j in Im_1, e.g.,

$$w_{ij}^I=e^{-\tilde{\beta}|I_i-I_j|}, \tag{19}$$

where I_i and I_j denote the intensities of the two neighboring pixels, and $\tilde{\beta}$ is a positive constant, and a measure w_{ij}^M reflecting the difference in the motion profiles associated to the two pixels (3).

At each coarsening step, we first determine the next coarser graph using the weighted aggregation procedure. This will produce a graph that is roughly equivalent to the finer level graph, with weights inherited from the previous level. We then modify the weights in the new graph to incorporate coarser measures of differences between neighboring aggregates. Specifically, for every two aggregates we multiply these weights by the a measure of difference between their average intensities and possibly their variance (of the form similar to (7)), and likewise by a measure reflecting the difference between their motion profiles (3), and at higher scales the difference between their affine transformations and fundamental matrices.

4. Experiments

We implemented the combined algorithm and applied it to a collection of image pairs. (The image pairs can be seen in motion in the supplementary file.) We set the parameters around the following values: $\alpha = 10$, $\beta = 4$, $\tilde{\beta} = 7$. The motion profile distance d_{profile} was evaluated from the finest scale to scale 4. The affine distance d_{affine} was evaluated from a scale 5, and $d_{\text{fundamental}}$ was applied at the two topmost levels. Our non-optimized implementation runs in less than 10 seconds on a 200×250 on a Pentium 4 PC. To demonstrate the handling of motion cues alone we first applied the algorithm to several pairs of images containing a moving sequence of random dots. Figure 1 shows a random dot sequence containing a pair of translating squares on a stationary background along with segmentation results (displayed by a color overlay on top of the leftmost picture) and motion vectors computed by our algorithm. In this and other examples we extracted the motion vectors identified in peaked aggregates at levels 4-5. The bar-peaked aggregates are not displayed in these images. A similar display is shown in Figure 2, which contains a central rotating square on a stationary background.

Figure 3 contains a random dot sequence of an internal sphere placed within a spherical room. (The "camera" is picturing the scene from within the larger scene.) The two sphere translate in 3D in different directions. The use of fundamental matrices was critical in this case for obtaining the outer sphere in one piece. This sphere was obtained in several pieces when we only used the affine transformation.

The rest of the figures show the results obtained with our method on real motion pairs. Figure 4 shows a car existing to a street with some camera motion. Using intensity alone resulted in attaching the dark top of the car with the background. Using also motion cues the car is segmented in one piece (although the wheels were incorrectly attached to the background). The figure further shows the epipolar lines computed with our method for the background segment. In the next example (Figure 5) the motion pair is blurred resulting in bleeding (the woman's face is segmented with the background) when only intensity cues are employed, where as motion cues caused the face and the hair to come out in one piece. In Figure 6 the arm of the octopus was connected in one piece despite a dark section in the middle mainly because of continuity in motion. Similar results are obtained in Figures 7 and 8.

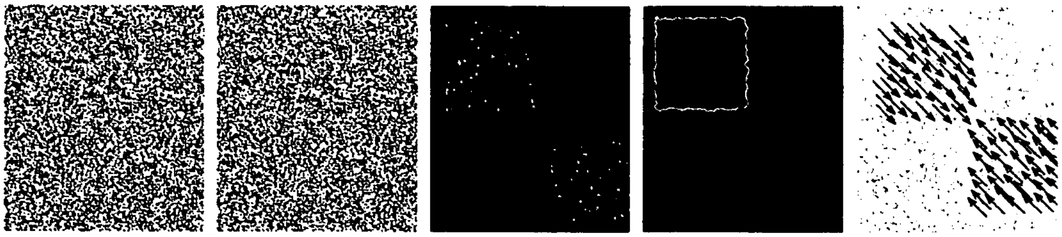


Figure 1: From left to right: a random dot pair containing two translating squares, a difference image, segmentation results obtained with our method, and motion vectors obtained from peaked aggregates at levels 4-5.

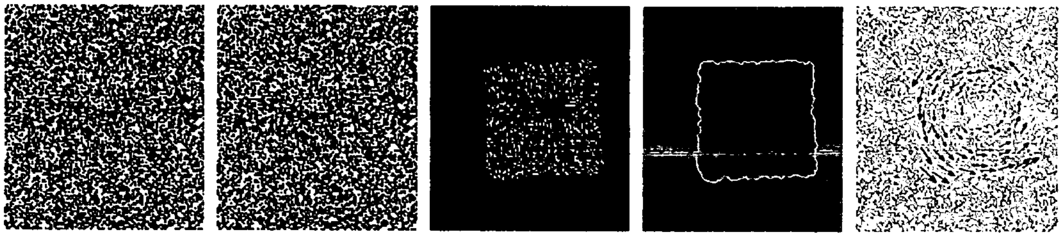


Figure 2: A random dot pair containing a central, rotating square, a difference image, segmentation results obtained with our method, and motion vectors.

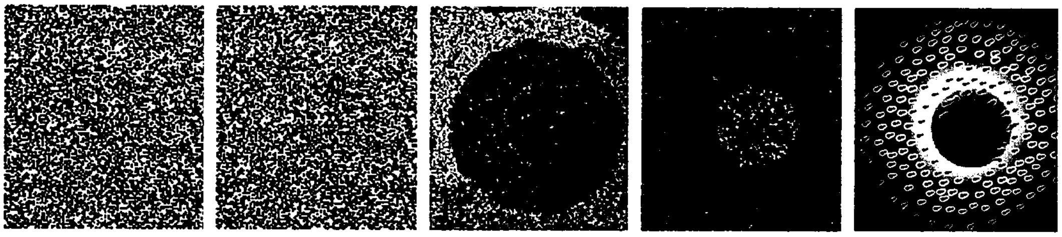
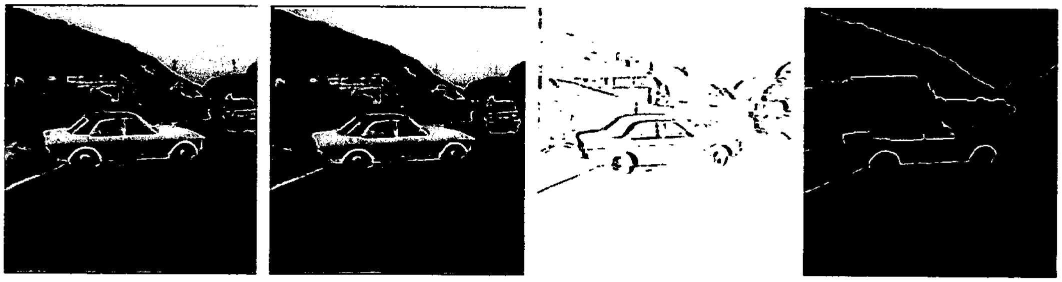


Figure 3: A random dot pair containing two spheres translating in 3D, results obtained by applying our method with affine transformation only, results obtained by applying a comparison of fundamental matrices at the coarsest levels, and motion vectors overlaid on a depth image (intensity proportional to distance from camera).



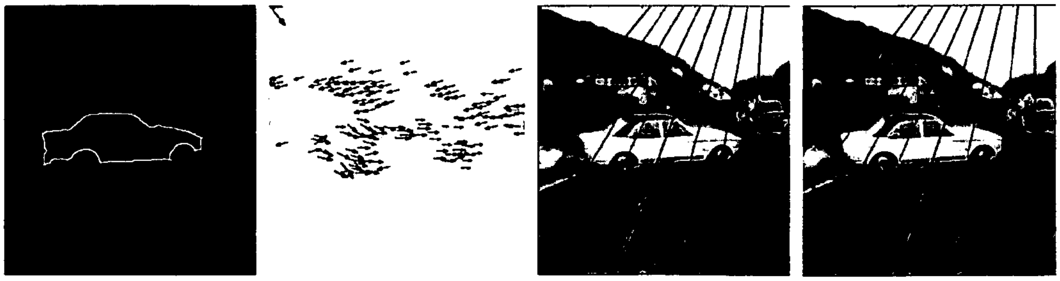


Figure 4: Top row: a motion pair, a difference image, results obtained by applying segmentation based on intensity cues alone. Bottom row: results obtained by applying our method with both motion and intensity cues, motion vectors, and epipolar lines computed for the background segment.

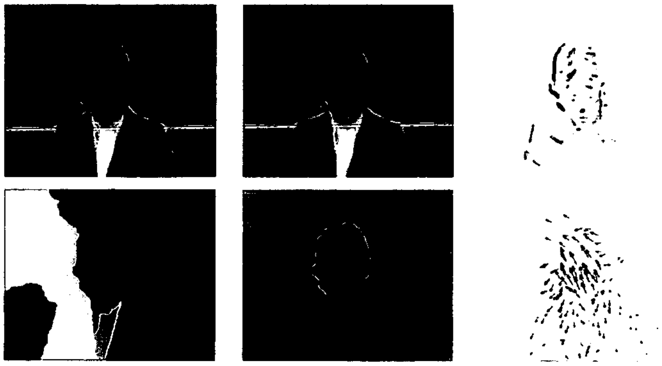


Figure 5: A motion pair, a difference image, results obtained by applying segmentation based on intensity cues alone, results obtained by applying our method with both motion and intensity cues, and motion vectors.

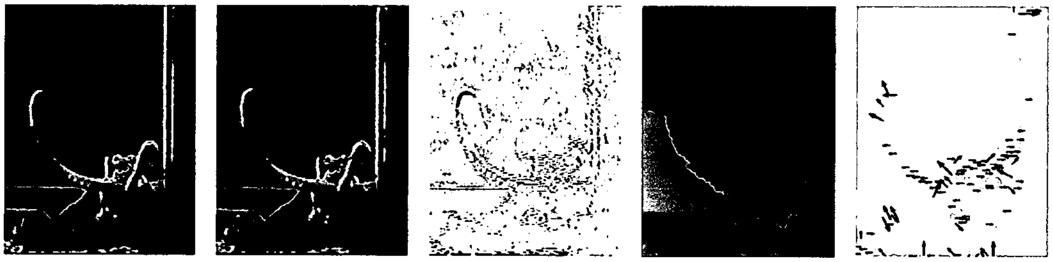


Figure 6: A motion pair, a difference image, results obtained by applying our segmentation method, and motion vectors.

5. Conclusion

We have presented an efficient multiscale algorithm for image segmentation that combines motion with intensity cues. The algorithm uses bottom-up processing to disambiguate motion measurements and determine an appropriate motion model for various regions in the image. We have demonstrated the algorithm by applying it to a variety of random dot and real motion pairs.

Our algorithm is related to several existing algorithms. We represent initial motion measurements in the form of motion profiles and apply a graph algorithm to find minimal cuts in the graph (as in [15]) using the algorithm proposed in [14]. In addition, similar to layer approaches [19,21,1,20] our algorithm is composed of a sequence of clustering and re-estimation steps. However, unlike these methods, our method uses coarsening steps to disambiguate motion measurements and to adaptively select a motion model according to the amount of statistics available. Because each coarsening step reduces the number of clusters handled by the algorithm the algorithm is linear in the number of pixels in the image.

Our method can be easily extended in several ways. First, it can be extended to handle sequences composed of three or more frames, e.g., by applying the same method on a 3D graph which represents the pixels in the entire sequence. Motion measurements at all levels can be more robustly computed by applying robust estimation techniques. A top-down scan of the pyramid can be applied to obtain a dense motion field. The method can be used to recover the 3D structure of a scene by using the motion vectors in conjunction with the epipolar constraints to determine disparities. Another straightforward extension is to incorporate 2D homographies as a motion model to correctly segment planes and to identify the background in the case of camera rotation. The method already accounts for some nonrigid deformations by allowing for stretch and skew in the affine model and by clustering neighboring aggregates when these have similar motion (see, e.g., the arm of the octopus, which is moving non-rigidly, yet it is segmented in one piece). More complex deformation models can also be incorporated in the method. Also of importance is to handle the "edge assignment" problem in order to determine for boundary edges to which segment their motion is relevant. This is important particularly in uniform regions when almost no supporting features exist in addition to the boundary edges. The use of 1×1 windows instead of 3×3 to initialize the motion profiles can alleviate this problem. Finally, learning approaches can be incorporated to automatically set the parameters of this process.

References

[1] S. Ayer, H.S. Sawhney, Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding, *ICCV*: 777-784, 1995.

[2] M.J. Black, A. Jepson, Estimating optical flow in segmented images using variable-order parametric models with local deformations, *PAMI*, (10): 972–986, 1996.

[3] T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, *ECCV*: 25—36, 2004.

[4] J. Costeira and T. Kanade, A Multi-body Factorization Method for Motion Analysis, *IJCV* (3): 159–179, 1998.

[5] D. Cremers, A Variational Framework for Image Segmentation Combining Motion Estimation and Shape Regularization, *CVPR* I: 53–58, 2003.

[6] C. Fowlkes, S. Belongie, F. Chung, J. Malik. Spectral Grouping using the nystr?m method", *PAMI*, (2): 214–225, 2004.

[7] M. Galun, E. Sharon, R. Basri, A. Brandt, Texture segmentation by multiscale aggregation of filter responses and shape elements, *ICCV*: 716–723, 2003.

[8] C.W. Gear, Multibody grouping from motion images, *IJCV*, (29): 133—

150, 1998.

[9] A. Gruber, Y. Weiss, Multibody factorization with uncertainty and missing data using the EM algorithm, *CVPR*, : 707–714, 2004.

[10] R.I. Hartley, In Defense of the Eight-Point Algorithm, *PAMI*, (6): 580–593, 1997.

[11] K. Kanatani, Evaluation and selection of models for motion segmentation. *ECCV*: 335–349, 2002.

[12] M. Pawan Kumar, P.H.S. Torr, A. Zisserman, Learning Layered Pictorial Structures from Video

[13] M. Irani, B. Rousso, S. Peleg, Detecting and tracking multiple moving objects using temporal integration, *ECCV*: 282—287, 1992.

[14] E. Sharon, A. Brandt, R. Basri, Segmentation and boundary detection using multiscale intensity measurements, *CVPR*, I:469–476, 2001.

[15] J. Shi, J. Malik, Motion segmentation and tracking using normalized cuts" *ICCV*: 1154–1160, 1998.

[16] P.H.S. Torr, A.W. Fitzgibbon, A. Zisserman, Maintaining Multiple Motion Model Hypotheses Over Many Views to Recover Matching and Structure, *ICCV*: 485–491 1998.

[17] W.S. Tong, C.K. Tang, G. Medioni, Simultaneous Epipolar Geometry Estimation and Motion Segmentation by 4D Tensor Voting in Joint Image Space, *PAMI*, (9), 1167–1184, 2004.

[18] P.H.S. Torr, A. Zisserman, S. Maybank, Robust Detection of Degenerate Configurations for the Fundamental Matrix, *CVIU*, (3): 312–333, 1998.

[19] J.Y.A. Wang, E.H. Adelson, Representing Moving Images with Layers. *IEEE Trans. on Image Processing*, (5):625–638, September 1994.

[20] Y. Weiss, Smoothness in Layers: Motion segmentation using nonparametric mixture estimation. *CVPR*, 520–527, 1997.

[21] Y. Weiss, E.H. Adelson, A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. *CVPR*, 321–326, 1996.

[22] L. Zelnik-Manor, M. Machline, M. Irani, Multi-Body factorization with uncertainty: revisiting motion consistency, IJCV, 2004.

WHAT IS CLAIMED IS:

1. A method for finding correspondence between portions of two images comprising the steps of
   a) subjecting the two images to segmentation by weighted aggregation employing a series of coarsening in successively coarser levels,
   b) modifying the weights in each successive level to incorporate coarser measures of difference between neighboring aggregates based on a measure of difference between their average intensities and by a measure reflecting their motion profiles, and
   c) recovering at the highest level a representation of the correspondence between the portion of the two images.

2. A method according to claim 1 wherein the weight is determined according to

$$w_{ij}^{I} = e^{-\tilde{\beta}|I_i - I_j|},$$

where I_i and I_j denote the intensities of the two neighboring pixels, and $\tilde{\beta}$ is a positive constant, and a measure w_{ij}^{M} reflecting the difference in the motion profiles associated to the two pixels.

3. Apparatus for finding correspondence between portions of two images comprising
   a. means for subjecting the two images to segmentation by weighted aggregation employing a series of coarsening in successively coarser levels,
   b. means for modifying the weights in each successive level to incorporate coarser measures of difference between neighboring aggregates based on a measure of difference between their average intensities and by a measure reflecting their motion profiles, and
   c. means for recovering at the highest level a representation of the correspondence between the portion of the two images.

4. Apparatus according to claim 3 further including:

means for determining the weight according to

$$w_{ij}^{I} = e^{-\tilde{\beta}|I_i - I_j|},$$

where I_i and I_j denote the intensities of the two neighboring pixels, and $\tilde{\beta}$ is a positive constant, and a measure w_{ij}^{M} reflecting the difference in the motion profiles associated to the two pixels.

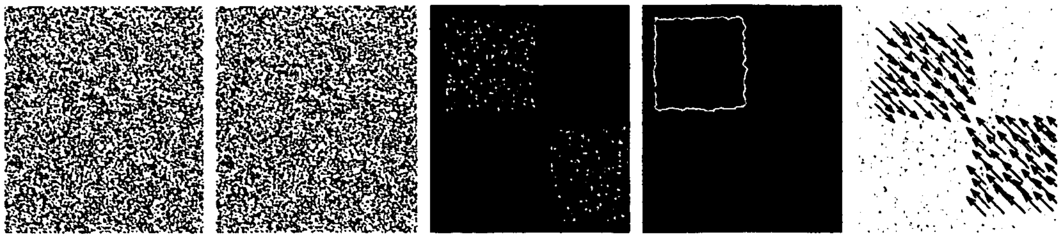


Figure 1: From left to right: a random dot pair containing two translating squares, a difference image, segmentation results obtained with our method, and motion vectors obtained from peaked aggregates at levels 4-5.

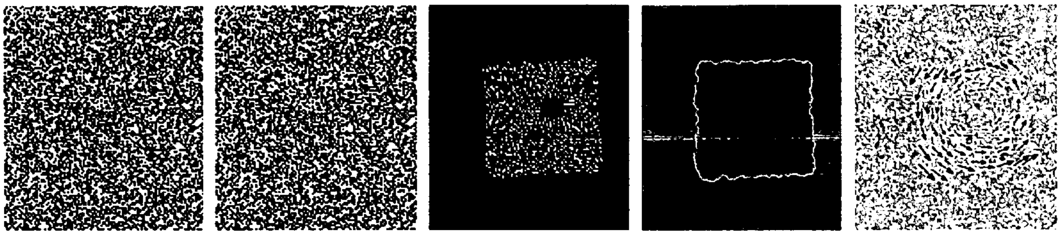


Figure 2: A random dot pair containing a central, rotating square, a difference image, segmentation results obtained with our method, and motion vectors.

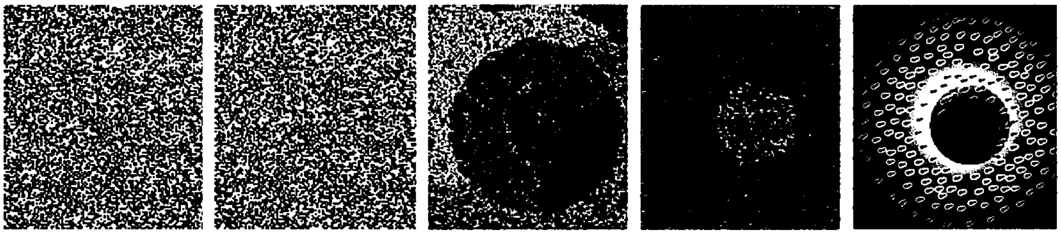


Figure 3: A random dot pair containing two spheres translating in 3D, results obtained by applying our method with affine transformation only, results obtained by applying a comparison of fundamental matrices at the coarsest levels, and motion vectors overlaid on a depth image (intensity proportional to distance from camera).

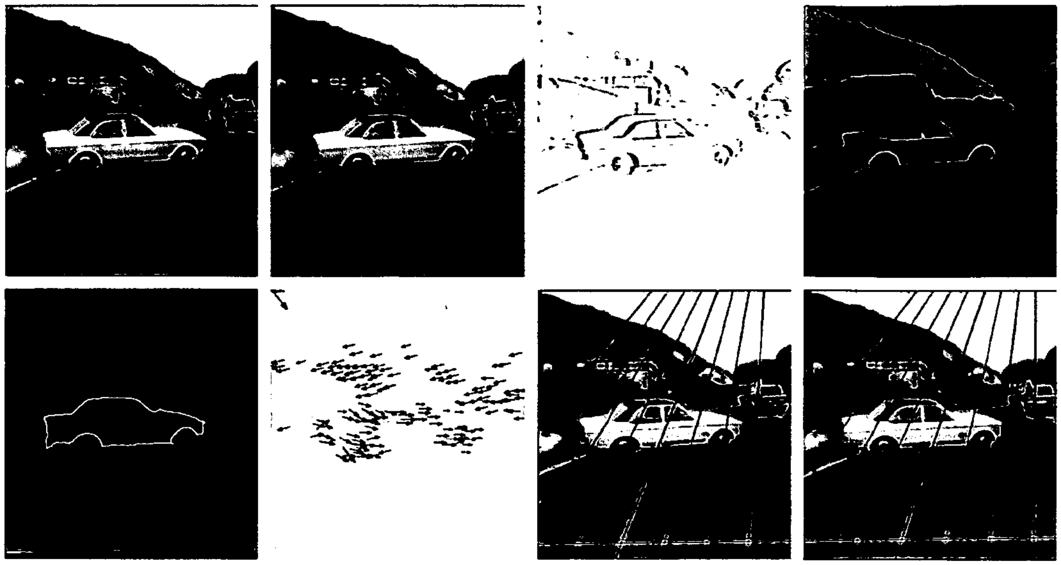


Figure 4: Top row: a motion pair, a difference image, results obtained by applying segmentation based on intensity cues alone. Bottom row: results obtained by applying our method with both motion and intensity cues, motion vectors, and epipolar lines computed for the background segment.

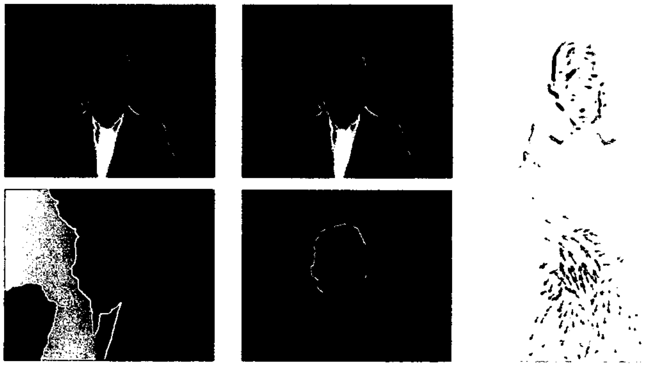


Figure 5: A motion pair, a difference image, results obtained by applying segmentation based on intensity cues alone, results obtained by applying our method with both motion and intensity cues, and motion vectors.

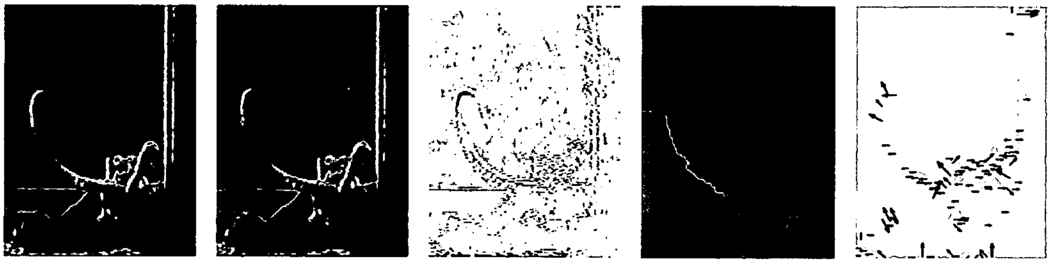


Figure 6: A motion pair, a difference image, results obtained by applying our segmentation method, and motion vectors.